

PENERAPAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK SEGMENTASI PELANGGAN POTENSIAL DI JEGER JERSEY INDONESIA

Adam Mukharil Bachtiar¹, Dian Dharmayanti², Riky Lutfi Hamzah³

^{1,2,3}Teknik Informatika, Universitas Komputer Indonesia,

Jalan Dipatiukur No.112-116, Bandung, 40132, Indonesia

Email: adam@email.unikom.ac.id¹, dian.dharmayanti@email.unikom.ac.id², rilutham@gmail.com³

ABSTRAK

Tim-tim *Indonesia Super League* (ISL) selalu mengeluarkan *jersey* terbaru mereka setiap musim kompetisi baru akan dimulai. Momen tersebut memberikan kesempatan bagi Jeger Jersey Indonesia untuk membuat promosi penjualan *jersey* kepada pelanggan baru dan pelanggan yang sudah ada (*existing customers*). Aspek kunci dalam membuat promosi penjualan adalah identifikasi terhadap perilaku pembelian pelanggan untuk membentuk segmen pelanggan yang memiliki pola serupa. Namun saat ini segmentasi pelanggan tersebut belum dibentuk, sehingga promosi penjualan bagi setiap pelanggan pun belum ditetapkan. Metode *Hierarchical Agglomerative Clustering* digunakan untuk menyegmentasikan pelanggan ke dalam segmen-segmen yang terbentuk secara alami berdasarkan atribut-atribut data. Segmentasi diawali dengan merepresentasikan setiap objek pada data sebagai satu segmen, kemudian dilakukan perhitungan jarak (*distance measure*) antar segmen tersebut. Setelah itu, sepasang segmen yang memiliki jarak terdekat digabungkan menjadi sebuah segmen baru. Proses perhitungan jarak dan penggabungan sepasang segmen dilakukan secara iteratif hingga tersisa satu buah segmen yang berisi seluruh objek. Jumlah segmen yang diinginkan dapat ditentukan dengan memotong *dendrogram* pada suatu jarak tertentu. Hasil penelitian menunjukkan bahwa metode tersebut berhasil membentuk segmentasi pelanggan. Jumlah segmen yang ditentukan merupakan yang paling optimal dan setiap segmen memiliki karakteristik yang berbeda. Oleh karena itu, hasil penelitian ini membantu manajer *marketing* dalam menetapkan promosi penjualan yang tepat untuk setiap segmen pelanggan.

Kata Kunci: segmentasi pelanggan, *clustering*, Hierarchical Agglomerative Clustering, *knowledge discovery*, *data mining*.

1. PENDAHULUAN

Jeger Jersey Indonesia adalah *startup* yang bergerak dalam bidang penjualan *merchandise* tim sepakbola Indonesia. Media yang menjadi saluran utama pemasarannya adalah situs web dan media sosial. Salah satu teknik pemasaran yang saat ini diterapkan adalah dengan memasarkan produk kepada pelanggan sesuai dengan tim sepakbola favoritnya. Menurut manajer *marketing*, tren pemasaran *online* telah berubah dari era pemasaran massal dimana suatu produk menjangkau semua orang menjadi pemasaran yang terfragmentasi. Strategi yang perlu diterapkan saat ini adalah memasarkan produk yang tepat kepada pelanggan yang tepat.

Tim-tim *Indonesia Super League* (ISL) selalu mengeluarkan *jersey* terbaru mereka setiap musim kompetisi baru akan dimulai. Momen tersebut memberikan kesempatan bagi Jeger Jersey Indonesia untuk membuat promosi penjualan *jersey* kepada pelanggan baru dan pelanggan yang sudah ada (*existing customers*). Memasarkan produk kepada pelanggan yang sudah ada menjadi prioritas untuk meningkatkan penjualan produk. Hal tersebut dapat dicapai dengan menerapkan promosi penjualan yang tepat untuk setiap pelanggan. Manajer *marketing* menyatakan bahwa aspek kunci dalam membuat promosi penjualan adalah identifikasi terhadap perilaku pembelian pelanggan untuk membentuk segmen pelanggan yang memiliki pola serupa.

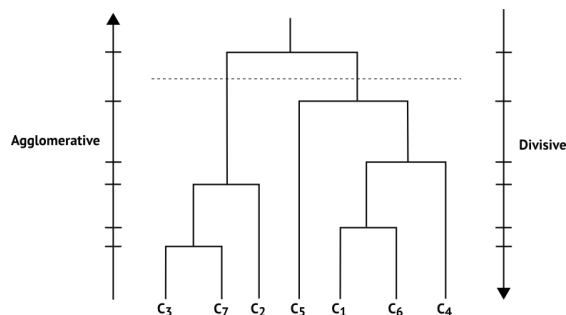
Teknik *clustering* dapat digunakan untuk menyegmentasikan pelanggan potensial berdasarkan kedekatan karakteristik antar pelanggan [1]. Salah satu metode yang dapat digunakan pada teknik *clustering* adalah metode *Hierarchical Agglomerative Clustering* [2]. Metode tersebut akan digunakan pada penelitian ini karena jumlah segmen yang akan dibentuk untuk segmentasi pelanggan belum diketahui. Selain itu, segmen-segmen pelanggan akan terbentuk secara alami berdasarkan atribut-atribut data pelanggan.

2. ISI PENELITIAN

2.1 Pemahaman Dasar Hierarchical Clustering

Hierarchical clustering merupakan metode *clustering* yang dapat melakukan pengelompokan objek pada data ke dalam sebuah hierarki [3]. Terdapat dua teknik pengelompokan objek pada *hierarchical clustering* yaitu secara *agglomerative (bottom-up)* dan *divisive (top-down)*. *Hierarchical agglomerative clustering* menggabungkan setiap objek hingga menjadi satu kelompok, sedangkan *hierarchical divisive clustering* memisahkan semua objek pada sebuah kelompok besar menjadi kelompok yang hanya memiliki satu objek. Ketika *agglomerative* telah menggabungkan dua kelompok maka mereka tidak dapat dipisahkan kembali, dan ketika *divisive* telah memisahkan dua objek maka mereka tidak dapat digabungkan kembali.

Hasil pengelompokan *hierarchical clustering* dapat direpresentasikan pada sebuah *dendrogram*. *Dendrogram* tersebut merupakan visualisasi struktur pengelompokan data dan dapat memberikan deskripsi yang informatif. Jumlah kelompok yang diinginkan dapat diperoleh dengan memotong *dendrogram* pada suatu jarak tertentu. Terdapat banyak metode dalam menentukan jumlah kelompok yang diinginkan dari sebuah *dendrogram* [4]. **Gambar 1** merupakan contoh *dendrogram* hasil pengelompokan data menggunakan metode *hierarchical clustering*.

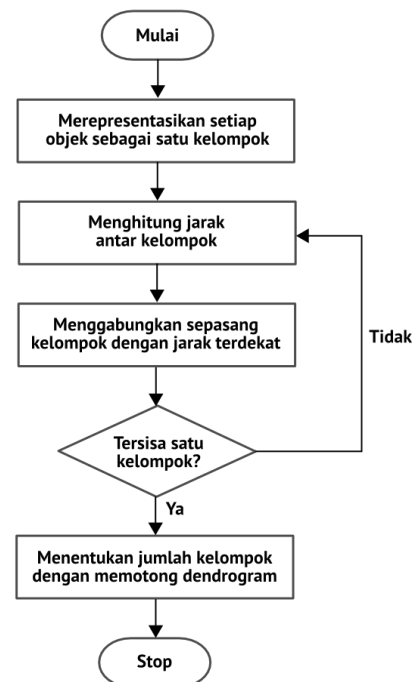


Gambar 1. *Dendrogram* hasil pengelompokan *Hierarchicalagglomerative Clustering* [4]

Pengelompokan data menggunakan metode *hierarchical agglomerative clustering* diawali dengan merepresentasikan setiap objek pada data sebagai satu kelompok, kemudian dilakukan perhitungan jarak (*distance measure*) antar kelompok tersebut. Setelah itu dua kelompok yang memiliki jarak terdekat digabungkan menjadi sebuah kelompok baru. Kemudian jarak antara kelompok yang baru dengan kelompok lain dihitung menggunakan salah satu metode perhitungan jarak antar kelompok (*single linkage, complete linkage, average linkage, dll.*). Proses perhitungan jarak dan penggabungan dua kelompok dilakukan secara iteratif hingga tersisa satu buah kelompok yang berisi seluruh objek. **Gambar 2**

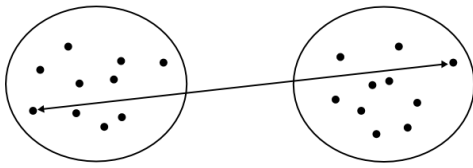
merupakan langkah-langkah dari metode *hierarchical agglomerative clustering*.

Perhitungan jarak antar kelompok menjadi faktor penting dalam metode ini. Proses tersebut dilakukan untuk meminimalkan jarak antar objek dalam satu kelompok (*intra-cluster distance*) dan memaksimalkan jarak antara objek dalam satu kelompok dengan objek dalam kelompok lain (*inter-cluster distance*) [4]. Dengan kata lain, perhitungan jarak antara dua kelompok perlu dilakukan untuk mengetahui kemiripan atau kedekatan antar kelompok tersebut.



Gambar 2. Langkah-langkah umum metode *Hierarchicalagglomerative Clustering*

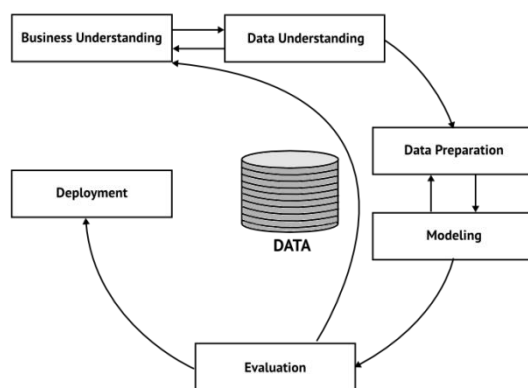
Salah satu metode yang dapat digunakan untuk menentukan jarak antar kelompok pada *hierarchical clustering* adalah metode *complete linkage*. Dengan metode *complete linkage*, jarak sepasang kelompok ditentukan oleh jarak terbesar antara dua objek pada kelompok yang dibandingkan [5]. *Complete linkage* membandingkan objek antar kelompok yang paling berbeda di setiap iterasi. Setelah perhitungan jarak dilakukan, dua kelompok yang memiliki jarak terkecil kemudian digabungkan. Gambaran perhitungan jarak metode *complete linkage* dapat dilihat pada Gambar 3.



Gambar 3 Gambaran perhitungan jarak metode *complete linkage*

2.2 Metodologi Penelitian

Metodologi penelitian yang digunakan pada penelitian ini mengacu ke standar CRISP DM yang memiliki enam langkah. Adapun gambaran alur proses *data mining* pada CRISP DM bisa dilihat pada Gambar 4.



Gambar 4. Model proses *data mining* pada CRISP DM [6]

Berikut adalah penjelasan dari langkah-langkah yang ada pada Gambar 4:

a. Pemahaman Tujuan Bisnis

Tahap ini fokus pada identifikasi tujuan proyek dan kebutuhan dari sudut pandang bisnis. Kemudian mendefinisikan masalah *data mining* dan menyusun rencana untuk mencapai tujuan proyek. Beberapa kegiatan pada tahap ini meliputi identifikasi tujuan bisnis, penilaian situasi, penentuan tujuan data mining, dan penyusunan rencana proyek yang akan dilakukan.

b. Pemahaman Data

Tahap ini dilakukan untuk mendeteksi bagian dari data yang mungkin mengandung informasi. Beberapa kegiatan pada tahap ini meliputi pengumpulan data awal, mendeskripsikan data yang diperoleh, mengeksplorasi data, dan melakukan verifikasi terhadap kualitas data [7].

c. Persiapan Data

Pada tahap ini dilakukan *data preprocessing* agar data dapat digunakan untuk proses pembentukan model. Beberapa kegiatan pada tahap

ini meliputi pemilihan data, pembersihan data, pembangunan data, mengintegrasikan data dari berbagai sumber, dan mentransformasikan data.

d. Pembentukan model

Tahap pembentukan model merupakan tahap utama pada proses *data mining*. Pada tahap ini diterapkan teknik pembentukan model pada data yang telah disiapkan. Beberapa kegiatan pada tahap ini meliputi pemilihan teknik pembentukan model yang sesuai dengan data, menjelaskan prosedur teknik pembentukan model yang digunakan, menerapkan teknik pembentukan model, dan menilai model yang dihasilkan.

e. Evaluation

Setelah model terbentuk, perlu dilakukan evaluasi terhadap langkah-langkah yang dilakukan sebelumnya. Hal tersebut dilakukan untuk memastikan model sesuai dengan tujuan bisnis yang ditetapkan. Beberapa kegiatan pada tahap ini meliputi evaluasi model yang dihasilkan, pengujian ulang proses-proses yang dilakukan, dan penentuan keputusan penggunaan hasil *data mining*.

f. Pengembangan

Tahap ini merupakan implementasi hasil dari proses *data mining*. Beberapa kegiatan pada tahap ini meliputi penentuan rencana penerapan hasil *data mining*, penentuan rencana pengawasan dan pemeliharaan, pembuatan laporan akhir, dan melakukan ulasan terhadap proyek yang telah dilakukan.

Pada penelitian ini tahapan pengembangan dilakukan dengan cara mengembangkan perangkat lunak guna menyampaikan pengetahuan yang terbentuk dari hasil penerapan algoritma *data mining*.

2.3 Hasil dan Pembahasan

Hasil dan pembahasan pada penelitian ini didasarkan pada model proses CRISP DM. Berikut adalah hasil dan pembahasan dari tiap tahapan:

a. Pemahaman Tujuan Bisnis

Pada langkah pertama dilakukan proses pemahaman bisnis dari Jeger Jersey agar pengetahuan yang didapat dari proses *data mining* berguna dan tepat sasaran. Fokus dari Jeger Jersey Indonesia saat ini adalah melakukan promosi penjualan kepada pelanggan yang sudah ada (*existing customers*). Pelanggan tersebut perlu diberikan pelayanan khusus untuk menjaga kepuasan dan loyalitas mereka. Pelanggan diharapkan akan melakukan pembelian kembali dan membeli produk *jersey* lebih banyak, sehingga dapat meningkatkan keuntungan dari penjualan produk *jersey*. Oleh karena itu, perlu diterapkan promosi penjualan yang sesuai dengan karakteristik setiap pelanggan.

Segmentasi pelanggan dapat dibentuk untuk mengetahui karakteristik dari setiap kelompok pelanggan. Segmentasi tersebut dapat dilakukan menggunakan salah satu teknik *data mining* yaitu metode *clustering* [8]. Hasil segmentasi diharapkan memiliki kualitas yang baik, karakteristik pelanggan pada sebuah segmen harus dapat dibedakan dengan pelanggan pada segmen lain.

Dalam penelitian ini, proses analisis data akan menggunakan bantuan perangkat lunak kode sumber terbuka (*open source software*). Penggunaan perangkat lunak *open source* sangat disarankan karena memiliki dokumentasi yang lengkap dan terbuka, mudah untuk dikustomisasi, dan kemudahan berkolaborasi dengan komunitas. Pengolahan data, pembentukan model, dan pengujian model akan dilakukan menggunakan bahasa pemrograman R.

b. Pemahaman Data

Data yang diperoleh adalah data pelanggan yang melakukan pembelian pada tahun 2013-2014. Data tersebut merupakan data pelanggan yang membeli berbagai produk seperti *jersey*, tas, jaket, dan lain-lain. Namun, data yang akan digunakan hanya data pelanggan yang melakukan pembelian produk *jersey*. Hal tersebut disesuaikan dengan tujuan bisnis yaitu meningkatkan penjualan produk *jersey*. Data tersebut diperoleh dari manajer *marketing* atas izin dari *owner* Jeger Jersey Indonesia.

Data pelanggan yang diperoleh disimpan pada sebuah berkas berekstensi **.csv* (*comma-separated values*). Data tersebut memiliki 12 atribut yang terdiri dari 6 atribut identitas pelanggan dan 6 atribut mengenai perilaku pembelian pelanggan. Jumlah baris atau *record* pada data tersebut sebanyak 808 baris. **Tabel 1** menyajikan penjelasan atribut-atribut data pelanggan Jeger Jersey Indonesia.

Tabel 1. Penjelasan atribut data pelanggan

Atribut	Tipe Data	Keterangan
ID	String	Identitas unik setiap pelanggan.
Nama	String	Nama pelanggan yang melakukan transaksi.
Alamat	String	Alamat pelanggan yang melakukan transaksi.
Kota	String	Kota pelanggan yang melakukan transaksi.
Provinsi	String	Provinsi pelanggan yang melakukan transaksi.
Telepon	String	Nomor telepon pelanggan yang melakukan transaksi.
Item	String	Jenis-jenis produk yang dibeli oleh pelanggan.
Custom Name?	Boolean	Apakah pelanggan membeli <i>custom name</i> ?

Atribut	Tipe Data	Keterangan
Jumlah Item	Numerik	Jumlah produk yang dibeli oleh pelanggan.
Jumlah Transaksi	Numerik	Jumlah transaksi yang dilakukan oleh pelanggan.
Total Transaksi	Numerik	Total transaksi seluruhnya yang dibayar oleh pelanggan.
Metode Pembayaran	String	Metode pembayaran yang digunakan oleh pelanggan.

Data pelanggan yang telah dideskripsikan perlu dieksplorasi secara menyeluruh untuk mengetahui kualitasnya. Eksplorasi tersebut meliputi kelengkapan data, validasi data, dan masalah-masalah lain yang terdapat pada data. Berikut merupakan hasil verifikasi kualitas data pelanggan:

- 1) Atribut *item* mengandung data teks bebas (*free text entries*).
- 2) Pada atribut *item*, banyak terdapat nilai yang berbeda tetapi memiliki makna yang sama (*inconsistent*).
- 3) Terdapat *missing value* pada atribut *item*, *custom name*, dan *jumlah item*.
- 4) Atribut *jumlah item* dan *jumlah transaksi* memiliki nilai yang bersifat *outliers*.

c. Persiapan Data

Atribut yang akan digunakan untuk proses pembentukan model adalah atribut *ID*, *item*, *custom name*, *jumlah item*, dan *jumlah transaksi*. Atribut-atribut tersebut menggambarkan perilaku pembelian pelanggan. Sedangkan *record* yang dipilih adalah semua pelanggan yang membeli produk *jersey*. *Record* tersebut dipilih sesuai dengan tujuan bisnis dari Jeger Jersey Indonesia yaitu meningkatkan penjualan produk *jersey*.

1) Pembersihan data

Proses pembersihan data dilakukan untuk memperbaiki kualitas data yang diperoleh. Pembersihan data ini mengacu pada verifikasi kualitas data yang sebelumnya telah dilakukan. Beberapa hal yang dilakukan pada proses pembersihan data adalah menangani data yang tidak konsisten, *missing value* dan *outliers*.

Atribut *item* yang mengandung data teks bebas (*free text entries*) akan dikodekan atau dikonversikan ke dalam data numerik. Setiap jenis produk *jersey* memiliki kode yang berbeda-beda. Hal tersebut dilakukan agar data pada atribut *item* menjadi konsisten. Kemudian *record* yang memiliki *missing value* pada atribut *item* akan dihapus, karena atribut *item* berpengaruh pada atribut *custom name* dan *jumlah item*.

Kemudian *outliers* pada atribut *jumlah item* dan *jumlah transaksi* akan tetap diproses karena data *outliers* tersebut merupakan data yang valid. Proses *smoothing* akan dilakukan terhadap data

pada atribut-atribut tersebut. Salah satu metode yang dapat digunakan untuk *smoothing* data adalah metode *binning* [9]. *Binning* merupakan metode yang digunakan untuk membagi sekumpulan nilai numerik ke dalam beberapa partisi (*bin*). Dengan menggunakan teknik ini, setiap nilai pada atribut akan didistribusikan ke dalam beberapa *bin* yang sudah ditentukan. **Tabel 2** merupakan contoh data pelanggan yang sudah dibersihkan.

Tabel 2. Penjelasan atribut data pelanggan

ID	item	custom_name	Jumlah_item	Jumlah_transaksi
JJ-0001	1	1	1	1
JJ-0005	10, 8, 3	0	≥ 3	≥ 2
JJ-0007	8	1	1	1
JJ-0008	5, 8	1	≥ 3	≥ 2
JJ-0011	11	1	1	1
JJ-0014	13	0	1	1
JJ-0026	3, 6	1	2	1
JJ-0024	8, 3, 6	1	≥ 3	≥ 2
JJ-0119	13, 8, 3, 1	0	≥ 3	≥ 2

2) Pembangunan Data

Tahap pembangunan data digunakan untuk membentuk atribut baru (*attribute construction*). Beberapa atribut seperti *item*, *jumlah_item*, dan *jumlah_transaksi* akan diturunkan menjadi beberapa atribut turunan dengan tipe data biner asimetris [10]. Hal tersebut dilakukan agar semua atribut yang digunakan untuk pembentukan model memiliki tipe data yang sama, sehingga akan memudahkan perhitungan kedekatan jarak antar pelanggan. **Tabel 3** merupakan contoh penurunan atribut pada atribut *jumlah_item*.

Tabel 3. Hasil penurunan atribut jumlah item

ID	Jumlah_item_1	Jumlah_item_2	Jumlah_item_3
JJ-0001	1	0	0
JJ-0005	0	0	1
JJ-0007	1	0	0
JJ-0008	0	0	1
JJ-0011	1	0	0
JJ-0014	1	0	0
JJ-0026	0	1	0
JJ-0024	0	0	1
JJ-0119	0	0	1

3) Penyusunan Data

Proses penyusunan data merupakan langkah terakhir dari tahap persiapan data. Pada tahap ini, dilakukan perubahan sintaksis pada data yang akan digunakan pada saat pembentukan model. Selain itu, urutan dari atribut juga harus diperhatikan. Atribut *ID* digunakan sebagai label, sedangkan 19 atribut selanjutnya merupakan atribut yang akan

digunakan untuk pembentukan model. Jumlah *record* yang digunakan sebanyak 773 pelanggan.

d. Pembentukan model

Pada tahapan ini dilakukan penerapan algoritma *data mining* pada kasus segmentasi pelanggan di Jeger Jersey. Berikut adalah gambaran proses penerapan algoritma Hierarchical agglomerative Clustering di penelitian ini:

1) Langkah pertama

Setiap pelanggan direpresentasikan sebagai satu kelompok. Pada data yang telah disiapkan terdapat sebanyak 773 pelanggan, sehingga akan terbentuk 773 kelompok. Setelah itu, dilakukan perhitungan jarak antara setiap kelompok. Metode yang digunakan untuk menghitungnya adalah metode *jaccard distance* [3]. Penggunaan metode tersebut didasarkan pada data yang memiliki tipe data biner asimetris.

2) Langkah kedua

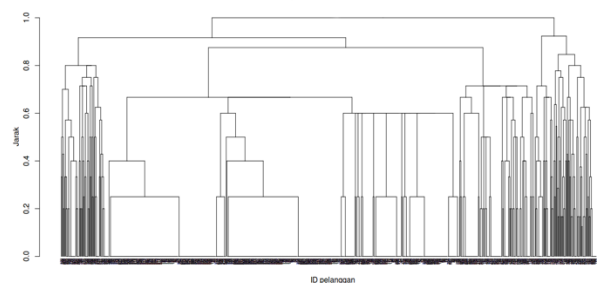
Setelah jarak antar kelompok pelanggan dihitung, langkah berikutnya adalah menggabungkan dua kelompok yang memiliki jarak terkecil (paling mirip).

3) Langkah ketiga

Kemudian dilakukan perhitungan jarak antara kelompok yang baru dengan kelompok yang lainnya. Metode yang digunakan adalah metode *complete linkage*. Jarak baru yang dihasilkan merupakan jarak terbesar antar anggota pada dua kelompok.

4) Langkah keempat

Ulangi langkah kedua dan ketiga sampai tersisa satu kelompok (kumpulan semua pelanggan). Pengulangan akan berhenti pada iterasi ke-772. **Gambar 5** merupakan hasil segmentasi pelanggan yang disajikan dalam sebuah *dendrogram*. Sumbu *x* merepresentasikan setiap pelanggan dan sumbu *y* merepresentasikan jarak antar pelanggan.

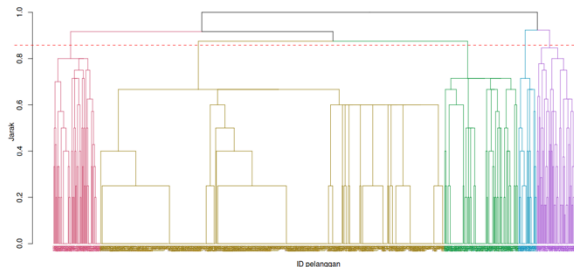


Gambar 5. Dendrogram hasil segmentasi pelanggan

5) Langkah kelima

Langkah terakhir pada metode *hierarchical agglomerative clustering* adalah menentukan

jumlah segmen yang diinginkan dengan memotong *dendrogram* pada jarak tertentu. Peneliti mencoba beberapa kemungkinan jumlah segmen yang dapat dibentuk sampai menghasilkan suatu pola atau pengetahuan. **Gambar 6** merupakan *dendrogram* yang dipotong pada suatu jarak, sehingga menghasilkan 5 segmen pelanggan.



Gambar 6. Dendrogram yang menghasilkan lima segmen pelanggan (dibedakan berdasarkan pewarnaan)

Tabel 4 merupakan statistik hasil segmentasi pelanggan yang terbentuk dari *dendrogram* pada Gambar 6.

Tabel 4. Statistik hasil segmentasi pelanggan di Jeger Jersey

Segmen ke-	Jumlah pelanggan	Jumlah Item	Jumlah Transaksi	Custom Name
1	69	≥ 3	1	42%
2	506	1	1	34%
3	109	2	1	29%
4	27	2	2	59%
5	62	≥ 3	≥ 2	77%

Berdasarkan analisis terhadap hasil segmentasi pelanggan pada Gambar 6, atribut *jumlah_item* dan *jumlah_transaksi* memiliki pengaruh yang besar terhadap proses segmentasi pelanggan. Pelanggan yang memiliki nilai *jumlah_item* dan *jumlah_transaksi* yang sama akan tergabung pada segmen yang sama. Jika diberikan data pelanggan baru, segmentasi pelanggan akan mudah dilakukan hanya dengan melihat kedua atribut tersebut. Model yang dihasilkan dapat membedakan karakteristik pelanggan pada suatu segmen dengan pelanggan pada segmen lain. Namun, masih perlu dilakukan evaluasi agar menghasilkan segmentasi pelanggan yang berkualitas.

Setelah lima segmen pelanggan terbentuk maka langkah berikutnya adalah melakukan pengukuran kualitas segmentasi pelanggan yang dihasilkan. Pengujian kualitas segmentasi akan dilakukan dengan menghitung rata-rata *silhouette coefficient* dari setiap objek pada data. Perhitungan *silhouette coefficient* dilakukan untuk mengetahui jumlah segmen yang paling optimal untuk data pelanggan yang dimiliki [9]. Jumlah segmen yang memiliki nilai rata-rata terbesar merupakan jumlah segmen

yang memiliki kualitas terbaik. Pengujian akan dilakukan beberapa kali dengan jumlah segmen yang berbeda. **Tabel 5** merupakan hasil perhitungan rata-rata *silhouette coefficient* pada beberapa jumlah segmen yang mungkin dibentuk.

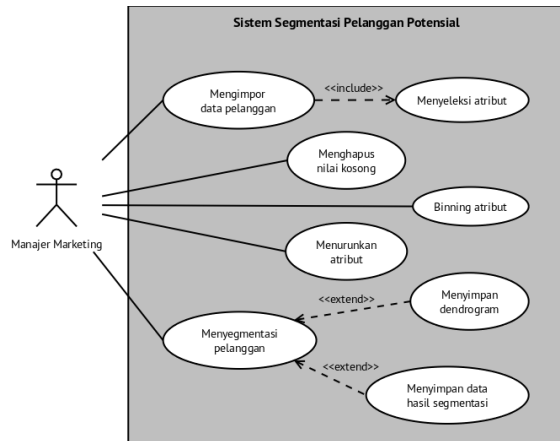
Tabel 5. Nilai rata-rata *silhouette coefficient* pada segmentasi pelanggan di Jeger Jersey

Jumlah segmen	Nilai rata-rata <i>silhouette coefficient</i>
2	0.3512
3	0.3196
4	0.2899
5	0.3753
6	0.3623
7	0.3283
8	0.3022
9	0.3003
10	0.3005
11	0.2958
12	0.2911
13	0.2899
14	0.2912
15	0.2846
16	0.2844
17	0.2625
18	0.2576
19	0.2525
20	0.2598

Berdasarkan nilai rata-rata *silhouette coefficient* pada Tabel 5, jumlah segmen 5 memiliki nilai rata-rata terbesar. Hal tersebut menggambarkan bahwa segmentasi dengan jumlah segmen 5 memiliki kualitas terbaik. Oleh karena itu, penentuan jumlah segmen yang dilakukan pada tahap pembentukan model sudah tepat. Selain itu, berdasarkan hasil wawancara dengan manajer *marketing* Jeger Jersey Indonesia, segmentasi pelanggan yang dihasilkan sudah sesuai dengan harapan. Manajer *marketing* dapat melihat dengan mudah karakteristik dari setiap segmen yang terbentuk. Oleh karena itu, manajer *marketing* dapat dengan mudah membuat strategi promosi penjualan yang sesuai dengan segmen-segmen tersebut.

e. Pengembangan

Pada tahapan pengembangan, proses di dalamnya dilakukan dengan menggunakan konsep SDLC (*Software Development Life Cycle*) pada umumnya. Perangkat lunak yang dibangun memiliki delapan fungsional yang tergambar pada use case diagram pada **Gambar 7**.



Gambar 7. Fungsional yang akan ada pada perangkat lunak segmentasi pelanggan di Jeger Jersey

Untuk bisa mengimplementasikan perangkat lunak yang dibangun, dibutuhkan lingkungan implementasi perangkat lunak seperti pada **Tabel 6**.

Tabel 6. Lingkungan implementasi perangkat lunak segmentasi pelanggan di Jeger Jersey

Jenis perangkat lunak	Spesifikasi
Sistem operasi	Fedora Workstation 21 64 bit
Kernel	Linux 3.18.7-200
Lingkungan desktop	GNOME 3.14
Bahasa pemrograman	Python 2.7.8
Library	PyQt 4, Pandas, Scipy, Scikit-Learn, Matplotlib
Teks editor	Visual studio code

Perangkat lunak yang dibangun memiliki beberapa antarmuka yang bisa diakses oleh manajer *marketing*. Berikut adalah beberapa antarmuka yang dimiliki oleh perangkat lunak yang dibangun:

1) Antarmuka pengolahan data

Antarmuka pengolahan data pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 8**.

Gambar 8. Antarmuka pengolahan data

2) Antarmuka visualisasi model

Antarmuka visualisasi model hasil segmentasi pelanggan (dendrogram) pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 9**.



Gambar 9. Antarmuka visualisasi model

3) Antarmuka hasil segmentasi

Antarmuka hasil segmentasi dalam bentuk tabel pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 10**.

Gambar 10. Antarmuka hasil segmentasi dalam bentuk tabel

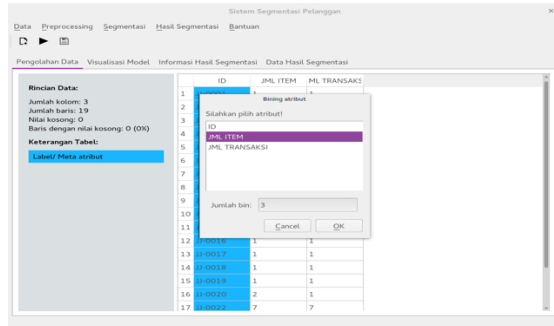
4) Antarmuka *import* data pelanggan

Antarmuka proses *import* data pelanggan pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 11**.

Gambar 11. Antarmuka *import* data pelanggan

5) Antarmuka *binning* data pelanggan

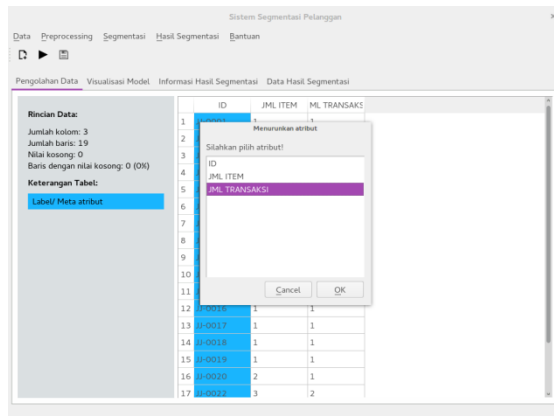
Antarmuka proses *binning* data pelanggan pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 12**.



Gambar 12. Antarmuka proses *binning* data pelanggan

6) Antarmuka penurunan atribut

Antarmuka proses penurunan atribut pada perangkat lunak segmentasi pelanggan di Jeger Jersey dapat dilihat pada **Gambar 13**.



Gambar 13 Antarmuka penurunan atribut

3. PENUTUP

Pada tahap ini dijelaskan mengenai kesimpulan dari penelitian yang telah dilakukan. Berdasarkan hasil pengujian kualitas segmentasi, jumlah segmen yang dibentuk merupakan jumlah segmen yang paling optimal. Penerapan beberapa proses pada tahap *data preprocessing* memiliki pengaruh besar pada hasil tersebut. Kemudian segmentasi pelanggan yang dihasilkan sudah sesuai dengan harapan manajer *marketing*. Selain itu, sistem segmentasi pelanggan yang dibangun membantu manajer *marketing* dalam menyegmentasikan pelanggan dan melihat karakteristik dari setiap segmen. Oleh karena itu, hasil penelitian ini membantu manajer *marketing* dalam menetapkan promosi penjualan yang tepat untuk setiap segmen pelanggan yang terbentuk.

Berikut merupakan beberapa saran yang dapat digunakan sebagai pertimbangan pada penelitian berikutnya:

- Menerapkan *text mining* untuk mengkonversikan data yang mengandung data teks bebas (*free text entries*) ke dalam data numerik.
- Menerapkan proses identifikasi *outliers* secara otomatis pada sistem segmentasi pelanggan. Hal tersebut akan mempermudah pengguna dalam menentukan penanganan terhadap *outliers*.
- Menambahkan beberapa jenis pengujian untuk menguji hasil segmentasi pelanggan. Hal tersebut perlu dilakukan agar segmentasi pelanggan yang dihasilkan memiliki kualitas yang sangat baik.

DAFTAR PUSTAKA

- [1] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert Syst. Appl.*, vol. 31, no. 1, pp. 101–107, Jul. 2006.
- [2] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
- [3] E. Sazonov, "Clustering (Xu, R. and Wunsch, D.C.; 2008) [Book review]," *IEEE Pulse*, vol. 1, no. 1, pp. 74–76, Jul. 2010.
- [4] B. Everitt and Wiley InterScience (Online service), *Cluster Analysis*. Wiley, 2011.
- [5] F. Gorunescu, *Data mining: concepts, models and techniques*. Springer, 2011.
- [6] SPSS, *CRISP-DM 1.0 step by step data mining guide*. 2009.
- [7] A. M. Bachtiar and H. Laksmiwati, "Development of requirement elicitation model for prediction of student achievement in university," in *2014 2nd International Conference on Information and Communication Technology, ICoICT 2014*, 2014.
- [8] D. S. Rajagopal, "Customer Data Clustering using Data Mining Technique," Dec. 2011.
- [9] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, vol. 54, no. Second Edition. 2006.
- [10] S. P. Ćurić, M. Vranić, and D. Pinter, "Improvement of Hierarchical Clustering Results by Refinement of Variable Types and Distance Measures," *Automatika*, vol. 52, no. 4, pp. 353–364, Jan. 2011.